Vol 438|22/29 December 2005|doi:10.1038/nature04341 nature

ARTICLES

Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*

James E. Galagan[1], Sarah E. Calvo[1], Christina Cuomo[1], Li-Jun Ma[1], Jennifer R. Wortman[2], Serafim Batzoglou[3], Su-In Lee[3], Meray Baştürkmen[4], Christina C. Spevak[4], John Clutterbuck[5], Vladimir Kapitonov[6], Jerzy Jurka[6], Claudio Scazzocchio[7], Mark Farman[8], Jonathan Butler[1], Seth Purcell[1], Steve Harris[9], Gerhard H. Braus[10], Oliver Draht[10], Silke Busch[10], Christophe D'Enfert[11], Christiane Bouchier[12], Gustavo H. Goldman[13], Deborah Bell-Pedersen[14], Sam Griffiths-Jones[15], John H. Doonan[16], Jaehyuk Yu[17], Kay Vienken[18], Arnab Pain[15], Michael Freitag[19], Eric U. Selker[19], David B. Archer[20], Miguel Á. Peñalva[21], Berl R. Oakley[22], Michelle Momany[23], Toshihiro Tanaka[24], Toshitaka Kumagai[25], Kiyoshi Asai[25], Masayuki Machida[25], William C. Nierman[2,26], David W. Denning[27], Mark Caddick[28], Michael Hynes[29], Mathieu Paoletti[20], Reinhard Fischer[18], Bruce Miller[30], Paul Dyer[20], Matthew S. Sachs[4], Stephen A. Osmani[22] & Bruce W. Birren[1]

The aspergilli comprise a diverse group of filamentous fungi spanning over 200 million years of evolution. Here we report the genome sequence of the model organism *Aspergillus nidulans*, and a comparative study with *Aspergillus fumigatus*, a serious human pathogen, and *Aspergillus oryzae*, used in the production of sake, miso and soy sauce. Our analysis of genome structure provided a quantitative evaluation of forces driving long-term eukaryotic genome evolution. It also led to an experimentally validated model of mating-type locus evolution, suggesting the potential for sexual reproduction in *A. fumigatus* and *A. oryzae*. Our analysis of sequence conservation revealed over 5,000 non-coding regions actively conserved across all three species. Within these regions, we identified potential functional elements including a previously uncharacterized TPP riboswitch and motifs suggesting regulation in filamentous fungi by Puf family genes. We further obtained comparative and experimental evidence indicating widespread translational regulation by upstream open reading frames. These results enhance our understanding of these widely studied fungi as well as provide new insight into eukaryotic genome evolution and gene regulation.

The aspergilli are a ubiquitous group of filamentous fungi spanning over 200 million years of evolution. Among the over 185 aspergilli are several that have an impact on human health and society, including 20 human pathogens as well as beneficial species used to produce foodstuffs and industrial enzymes[1]. Within this genus, *A. nidulans* has a central role as a model organism. In contrast to most aspergilli, *A. nidulans* possesses a well-characterized sexual cycle and thus a well-developed genetics system. Half a century of *A. nidulans* research has advanced the study of eukaryotic cellular physiology, contributing to our understanding of metabolic regulation, development, cell cycle control, chromatin structure, cytoskeletal function, DNA repair, pH control, morphogenesis, mitochondrial DNA structure and human genetic diseases.

We present here the genome sequence for *A. nidulans*, and a comparative genomics study with two related aspergilli: *A. fumigatus*[2] and *A. oryzae*[3]. *A. fumigatus* is a life-threatening human pathogen, and

[1]The Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, Massachusetts 02142, USA. [2]The Institute for Genomic Research, Rockville, Maryland 20850, USA. [3]Department of Computer Science, Stanford University, Stanford, California 94305, USA. [4]Department of Environmental and Biomolecular Systems, Oregon Health & Science University, 20 000 NW Walker Road, Beaverton, Oregon 97006-8921, USA. [5]Division of Molecular Genetics, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G11 6NU, UK. [6]Genetic Information Research Institute, 1925 Landings Drive, Mountain View, California 94043, USA. [7]Institut de Génétique et Microbiologie, Institut Universitaire de France, Université Paris-Sud, UMR8621, 91405 Orsay Cedex, France. [8]Department of Plant Pathology, Plant Science Building, 1405 Veteran's Drive, University of Kentucky, Lexington, Kentucky 40546-0312, USA. [9]Plant Science Initiative and Department of Plant Pathology, University of Nebraska, Lincoln, Nebraska 68588, USA. [10]Department of Molecular Microbiology and Genetics, Institute for Microbiology and Genetics, Georg-August-University Gottingen, Grisebachstr. 8, D-37077 Gottingen, Germany. [11]Institut Pasteur, Unité Postulante Biologie et Pathogénicité Fongiques, INRA USC 2019, 75724 Paris Cedex 15, France. [12]Institut Pasteur, Génopole-PF1, 75724 Paris Cedex 15, France. [13]Faculdade de Ciencias Farmaceuticas de Ribeirao Preto, Universidade de Sao Paulo, Brazil. [14]Department of Biology, Texas A&M University, College Station, Texas 77843, USA. [15]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. [16]John Innes Centre, Colney Lane, Norwich NR4 7UH, UK. [17]Department of Food Microbiology and Toxicology, The University of Wisconsin-Madison, 1925 Willow Drive, Madison, Wisconsin 53706-1187, USA. [18]Max Planck Institute for terrestrial Microbiology, D-35043 Marburg and Institute for Applied Biosiences at the University of Karlsruhe, D-76187 Karlsruhe, Germany. [19]Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA. [20]School of Biology, University Park, University of Nottingham, Nottingham NG7 2RD, UK. [21]Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, Madrid 28040, Spain. [22]Department of Molecular Genetics, The Ohio State University, Columbus, Ohio 43210, USA. [23]The University of Georgia, Department of Plant Biology, 2502 Plant Sciences, Athens, Georgia 30602-7271, USA. [24]National Institute of Technology and Evaluation (NITE), 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan. [25]National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8566, Japan. [26]The George Washington University School of Medicine, Department of Biochemistry and Molecular Biology, 2300 Eye Street NW, Washington DC 20037, USA. [27]Schools of Medicine and Biological Sciences, The University of Manchester, Stopford Building, Manchester M23 9PL, UK. [28]Plant Science and Fungal Molecular Biology Research Group, School of Biological Sciences, Donnan Labs, The University of Liverpool, Liverpool L69 7ZD, UK. [29]Department of Genetics, University of Melbourne, Parkville, Victoria 3010, Australia. [30]Department of Microbiology, Molecular Biology and Biochemistry, University of Idaho, Moscow, Idaho 83844-3052, USA.

© 2005 Nature Publishing Group

A. oryzae is used in the production of sake, miso and soy sauce. *A. oryzae* and *A. fumigatus* lack known sexual cycles, and their study relies on *A. nidulans* as a genetic model. Our analysis of these organisms focused on the genomic bases of their differing physiologies, while investigating their common eukaryotic biology. Our results yield new insights into eukaryotic genome evolution, the evolution of mating-type loci, the potential for sexual reproduction in the two asexual species, and the role of conserved sequence elements in gene regulation.

Genome assembly and annotation

The genome sequence of *A. nidulans* was assembled from deep whole-genome shotgun (WGS) coverage obtained by paired-end sequencing from a variety of clone types (see Methods). An average of 13× sequence coverage was generated including ×3 coverage produced and provided by Monsanto. The Arachne package (http://www.broad.mit.edu/wga/) was used to assemble the sequence, and the resulting assembly consists of 248 sequence contigs with an N50 length of 282 kilobases (kb) (that is, 50% of all bases are contained in contigs of at least 282 kb). Contigs were assembled into 89 scaffolds with a total length of 30.06 megabases (Mb) (including gaps between contigs) and an N50 length of 2.44 Mb. A total of 28.5 Mb (95%) of the assembly was anchored to the *A. nidulans* genetic map[4,5] through meiotically mapped markers with sequence and markers located by haploidization or hybridization to electrophoretically separated chromosomes (see Supplementary Information). By comparison with previously published pulse-field gel electrophoresis data, we estimate that the assembly comprises 96.3% of the complete genome.

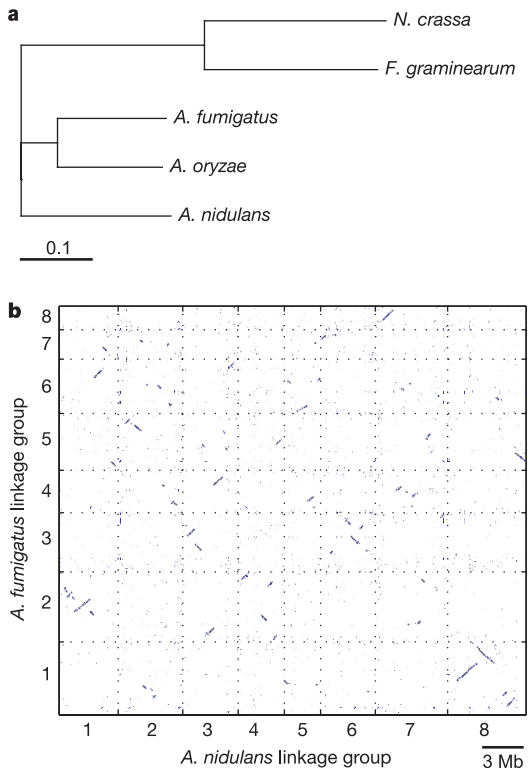

Figure 1 | Phylogenetic tree and representative dot plot. a, Phylogenetic tree showing the relationship between three *Aspergillus* species compared using *N. crassa* and *F. graminearum* as an outgroup. Branch lengths correspond to substitutions per site calculated using a maximum likelihood approach. An identical topology was predicted using *C. immitis* as an outgroup. **b**, Dot plot of *A. nidulans* (horizontal) and *A. fumigatus* (vertical) genomes. Axes represent the concatenation of all chromosomes for the corresponding genome. Gridlines indicate the boundaries between chromosomes and axis labels indicate chromosome number. Elements in the dot plot represent protein homology translated to genomic coordinates.

The assembly was annotated using the Calhoun system, as described in the Methods and Supplementary Information.

Phylogenetic relationship

Previous work based on large subunit rDNA data has led to a widely accepted phylogeny of the aspergilli in which *A. nidulans* and *A. oryzae* are more related to one another than *A. fumigatus*[6]. However, single gene phylogenies can contradict organismal phylogenies[7]. In principle, whole-genome data provide greater resolving power by allowing trees to be constructed based on concatenated sets of genes[7]. Using this approach to study the relationship of the three aspergilli, we find support for an alternative phylogeny[8] (Fig. 1a).

We established this relationship using a set of 3,034 predicted orthologues across the three aspergilli, *Neurospora crassa* and *Fusarium graminearum*. We constructed trees for 75 randomly selected sets of 20 concatenated genes[7], using the *N. crassa* and *F. graminearum* genes to root the trees (see Methods). All 75 cases produced the phylogeny shown in Fig. 1a in which *A. fumigatus* and *A. oryzae* are sister taxa and *A. nidulans* branches earlier. This phylogeny is further supported by 86% of trees built for each of the 3,034 orthologues individually. Consistent with this phylogeny, *A. fumigatus* has over twice as many genes with top Blast hits to *A. oryzae* than to *A. nidulans*, and *A. oryzae* has almost twice as many genes closer to *A. fumigatus* than *A. nidulans*. *A. nidulans* has roughly a similar number of top hits to *A. fumigatus* and *A. oryzae*. To confirm further the rooting of the tree, we repeated the analysis using predicted gene fragments (see Methods) from the genome sequence of *Coccidioides immitis* as an outgroup (which is closer to the aspergilli than *N. crassa* and *F. graminearum*). Ninety-four per cent (34 out of 36) of 50-gene phylogenies with *C. immitis* as the outgroup support the relationship of Fig. 1a, as do 60.8% (93 out of 153) of single gene phylogenies (only 21% support the rDNA phylogeny).

Overall genome and proteome comparison

Although in the same genus, the three aspergilli differ considerably in their genome sequences. Predicted orthologues shared by all three species (three-way orthologues) display an average of only 68% amino acid identity. *A. fumigatus* and *A. oryzae* share 70% identity, and each has 66–67% identity with *A. nidulans*. This protein identity is comparable to that between mammals and fish[9], which diverged ~450 million years ago. The three species also differ considerably in genome size (Table 1). The largest, *A. oryzae* (36 Mb), is 31% bigger than the smallest, *A. fumigatus* (28 Mb), and 24% bigger than *A. nidulans* (30 Mb). This difference seems to be due to an acquisition of sequence in *A. oryzae*[3] rather than loss in both *A. nidulans* and *A. fumigatus*. Finally, the genomes show extensive structural reorganization (Fig. 1b).

Conserved synteny and genome evolution

These three aspergilli provide an opportunity to study eukaryotic genome evolution over a divergence approaching the limit of conserved long-range synteny. To characterize pairwise conserved synteny, we used an algorithm based on hierarchical clustering that delineates regions of conserved synteny while also retaining information about internal micro-rearrangements (see Methods). Using this method, the majority (77–79%) of each genome assembly could be mapped to conserved syntenic blocks with at least one other genome (Table 2). Figure 2 shows a projection of the homologous blocks onto the chromosomes of *A. nidulans* and, contrasted with Fig. 1b, illustrates the considerable extent of conserved synteny despite extensive rearrangement.

The results of this analysis reveal two notable trends. First, large regions lacking detectable long-range synteny are readily apparent. As has been observed for mammals, nematodes and yeasts[10], repeats and subtelomeric sequences are associated with these heavily rearranged regions. This may have specific implications for fungi, as subtelomeric regions in the aspergilli are enriched for secondary

© 2005 Nature Publishing Group

Table 1 | Comparison of genome characteristics

Genome characteristic	*A. nidulans*	*A. fumigatus*	*A. oryzae*
General			
Assembly size (bp)	30,068,514	27,980,910	37,047,050
G+C (%)	50	49	48
Protein coding genes	9,541	9,926	14,063
Protein coding genes >100 amino acids	9,396	9,009	12,074
Predicted protein coding sequences >100 amino acids			
Coding (%)	50	49	45
Gene density (1 gene every *n* bp)	3,151	2,938	2,613
Median gene length (mean)	1,547 (1,868)	1,389 (1,644)	1,152 (1,414)
Average number of exons per gene	3.6	2.8	2.9

metabolite genes thought to have a role in niche adaptation and virulence. The rapid rearrangement of subtelomeric regions may facilitate the species-specific evolution of these genes (Supplementary Information).

The second notable trend is the distribution of lengths of unbroken regions between micro-rearrangements within pairwise syntenic blocks. The random breakage model of genome evolution predicts that such lengths should be exponentially distributed. Although the mean breakpoint lengths differ, in all three pairwise comparisons the distribution of lengths shows close agreement with the model prediction (Supplementary Information). It thus seems that syntenic blocks, comprising the majority of the *Aspergillus* chromosomes, are evolving in a manner consistent with random breakage.

For each pairwise comparison, the third *Aspergillus* genome allows the determination of rearrangements specific to each branch of the unrooted tree (see Methods). The results of this analysis provide a quantitative estimate of the different rearrangements contributing to long-term eukaryotic genome evolution (Fig. 3).

Structural evolution not correlated with molecular evolution

In vertebrates, nematodes and arthropods, it has been reported that the rates of structural evolution and nucleotide evolution are correlated[11–13]. However, our analysis of the *Aspergillus* genomes suggests that this expected correlation does not always hold for eukaryotes.

The data in Fig. 3 reveal a considerably higher overall rate of genome reorganization in the lineage of *A. oryzae* compared to *A. fumigatus*. Nearly all categories of disruption are at least twofold greater in *A. oryzae* relative to *A. fumigatus*. For example, *A. oryzae* displays a more than twofold greater rate of insertion than *A. fumigatus*. This is consistent with the larger genome size of *A. oryzae*[3]. Surprisingly, our analysis also indicates that chromosomal breaks are more common in *A. oryzae* than *A. fumigatus*. Although apparent intrachromosomal rearrangements could arise from successive inversion events, this cannot explain interchromosomal rearrangements. These interchromosomal breaks are also not the result of assembly error, as confirmed by optical mapping[3] and polymerase chain reaction (PCR) validation of eight predicted interchromosomal breaks.

In contrast, several measures indicate that the rates of amino acid evolution in predicted orthologues are similar between these two species. An examination of predicted three-way orthologues shows that the distribution of amino acid identity is roughly similar for both *A. oryzae* and *A. fumigatus* compared to *A. nidulans*, as are non-synonymous divergences (Supplementary Information). In addition, branch lengths predicted from phylogenetic trees (see above) indicate a comparable rate of substitution for both *A. oryzae* and *A. fumigatus*. Taken together, these data lead to the conclusion that structural and molecular evolution in the aspergilli is not correlated. A similar conclusion has been reached in the analysis of two microsporidian genomes, although in this case gene evolution seems to be accelerated relative to genome rearrangement[14]. Thus, large-scale and small-scale evolutionary processes in eukaryotes can operate at different relative rates in a species-specific manner.

Sex and the evolution of the mating-type loci

Unlike *A. nidulans*, which has a known sexual cycle, *A. fumigatus* and *A. oryzae* are only known to reproduce through asexual mitotic spores. We sought insight into the evolution of this apparent difference by comparing the three genomes. Our results, in conjunction with an accompanying paper and another study[2,15], suggest that both *A. fumigatus* and *A. oryzae* may be capable of sexual reproduction.

Sexual reproduction in ascomycete filamentous fungi is governed, in part, by two different mating-type genes that establish sexual compatibility: one gene encodes a protein with a high mobility group (HMG) domain, whereas the other encodes a protein with an alpha box domain. We refer to these genes here as the HMG and alpha mating-type genes, and to their chromosomal locations as MAT loci. Homothallic fungi typically possess both mating-type genes and are self-fertile. Heterothallic fungi possess only one mating-type gene and require a partner with a different mating-type gene. In heterothallics, the two mating-type genes typically occupy the same chromosomal location in different haploid genomes but lack sequence similarity, and are thus termed idiomorphs rather than alleles[16].

A. nidulans is known to be homothallic, and both HMG and alpha mating-type genes have been identified[17,18]. Our analysis confirmed that the HMG and alpha loci are unlinked, which is unusual although not unprecedented in homothallic fungi[19]. We identified a single HMG mating-type gene in *A. fumigatus*, as previously reported[20], and a single alpha mating-type gene in *A. oryzae*.

A comparison of all four MAT loci revealed extensive conserved

Table 2 | Characteristics of pairwise conserved synteny

Reference	Coverage (Mb) (percentage of reference)*				Maximum/mean block length (kb)‡		
	A. nidulans	*A. fumigatus*	*A. oryzae*	Either†	*A. nidulans*	*A. fumigatus*	*A. oryzae*
A. nidulans	–	20.5 (68)	20.4 (68)	21.6 (72)	–	175	114
A. fumigatus	20.4 (73)	–	20.7 (74)	21.5 (77)	2,429	–	168
A. oryzae	23.3 (63)	24.3 (66)	–	25.4 (69)	943	1,159	–

*Coverage of reference organism assembly by pairwise conserved syntenic blocks (>10 kb in length) to each target genome.
†Coverage of reference organism assembly by pairwise conserved syntenic blocks (>10 kb in length) in either other genome.
‡Upper right half shows mean blocks sizes and lower left half shows maximum sizes across all blocks using either organism as reference.

© 2005 Nature Publishing Group

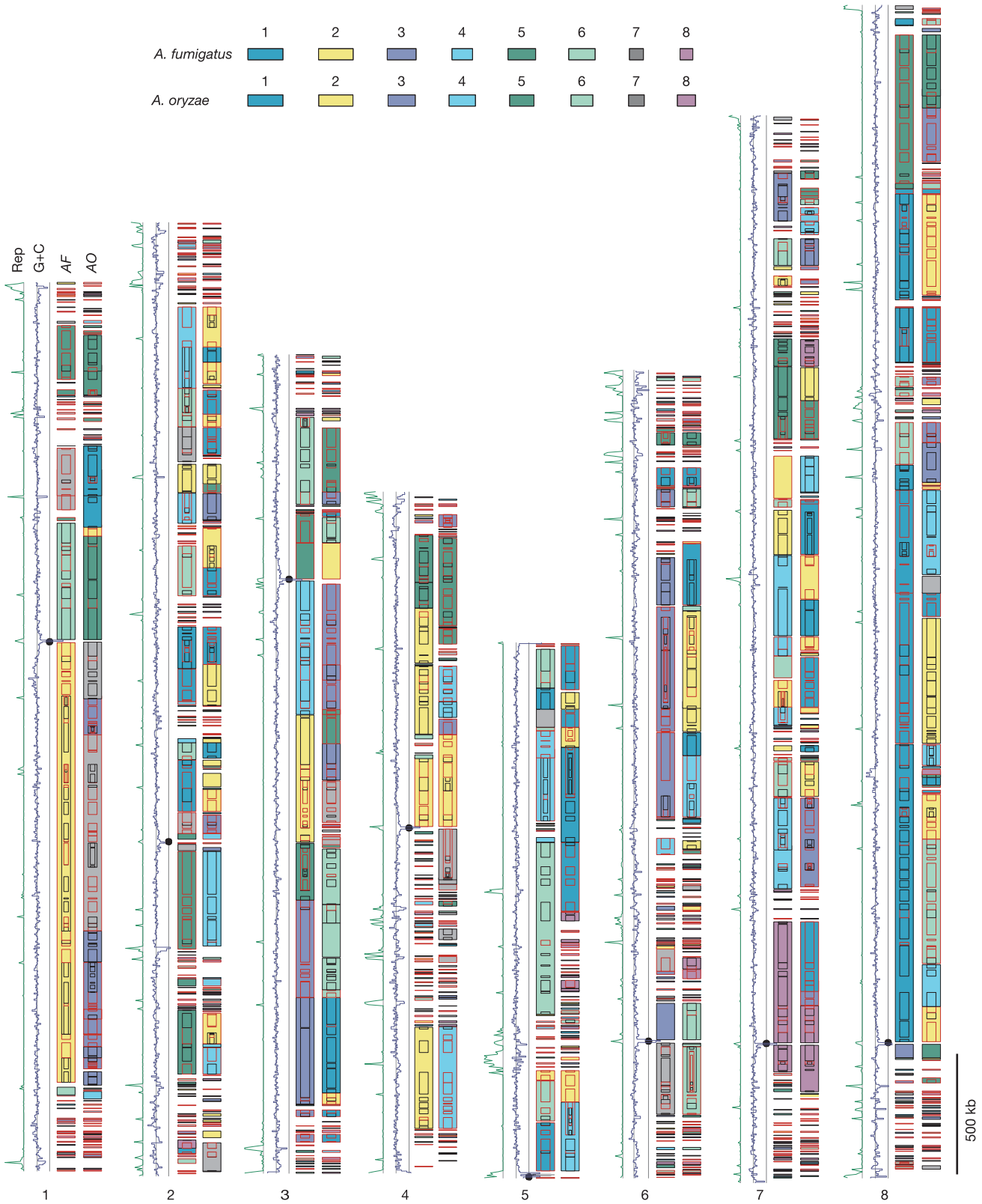

Figure 2 | *Aspergillus* comparative map. Conserved synteny between *A. nidulans* and *A. oryzae* and *A. fumigatus*. Syntenic regions are represented by two vertical columns of coloured blocks. The left and right columns represent syntenic blocks to *A. fumigatus* and *A. oryzae*, respectively, coloured by chromosome as indicated by the key. Nested blocks show synteny at finer resolutions. Blocks outlined in black are in opposite orientations in *A. nidulans* relative to those in red. Red blocks in black blocks (and vice versa) represent inversions. The green and purple lines display repeat density (Rep) and G+C content (G+C) in *A. nidulans*, both in 5-kb windows with increasing values to the left. Black circles represent centromeres.

© 2005 Nature Publishing Group

synteny (Fig. 4a). The *A. oryzae* alpha locus and the *A. fumigatus* HMG locus display conserved synteny over 1.7 Mb on either side of the mating-type genes. Within this region of conserved synteny, the two mating-type genes occupy nearly identical positions, although offset with different orientations. Furthermore, one flank of both the *A. fumigatus* and *A. oryzae* loci is syntenic with 409 kb of the *A. nidulans* HMG locus downstream region, whereas the other flank is syntenic with 34 kb of the *A. nidulans* alpha locus downstream region (Fig. 4a). The four loci also show conservation of a number of genes associated with MAT loci in other species including *N. crassa* or one of nine yeast species previously analysed[21].

Extending the analysis to 215 genes implicated in the fungal mating process, pheromone response, meiosis and fruiting body development revealed that every gene (except for the mating-type genes) that can be identified in *A. nidulans* is also present in both *A. fumigatus* and *A. oryzae* (Supplementary Information), including several genes for which the only known function is related to sexual reproduction.

Although sexual reproduction may have been lost very recently in both *A. fumigatus* and *A. oryzae*, providing one explanation for the residual presence of mating process genes, these data suggested the possibility that both *A. fumigatus* and *A. oryzae* may be capable of sexual reproduction. Moreover, the pattern of synteny among the four MAT loci leads to an evolutionary scenario for this hypothesis, as shown in Fig. 4b. According to this model, it is predicted that *A. oryzae* and *A. fumigatus* isolates exist with the opposite mating-type genes to those present in the strains that were sequenced. In addition, these opposite mating-type genes should be present at the identical locus, consistent with a heterothallic idiomorphic configuration.

As reported in detail in another study[15], these predictions have been experimentally verified. Using a PCR-based multiplex mating-type assay, isolates of both mating types of *A. fumigatus* and *A. oryzae* were identified. For both species, the opposite MAT locus from the complete genome was sequenced and demonstrated to have the idiomorphic organization predicted. Within the idiomorphic region the opposite mating-type genes appear to be offset with respect to one another, as predicted by our model. In addition, the *A. fumigatus* alpha MAT locus was found to contain a 360-base pair (bp) fragment of an HMG gene[15] neighbouring the idiomorphic region, suggesting that the transition from homothallism to heterothallism in the *A. oryzae* and *A. fumigatus* ancestor occurred by gene loss (Fig. 4b).

Although the model of Fig. 4b predicts a homothallic ancestor for all three species, it is possible that heterothallism was ancestral and a transition to homothallism occurred in the *A. nidulans* lineage. This would be consistent with data from *Cochliobolus* species for which heterothallism appears to be ancestral, and conversions to homothallism have been described[19]. However, two factors conflict with this scenario for the aspergilli. First, the offset positions of the mating-type genes within the idiomorphic regions of the *A. fumigatus* and *A. oryzae* MAT loci, and the apparent fragment of the HMG gene neighbouring the *A. fumigatus* alpha locus, are consistent with gene loss from a homothallic ancestor. Second, heterothallism in the aspergilli is rare[22,23]. Only three heterothallic aspergilli have been previously characterized, of which one, *A. heterothallicus*, groups in phylogenies with known homothallic species, suggesting a conversion to heterothallism in this case as well[22]. Mitotic, homothallic and heterothallic species are observed intermixed in several fungal lineages, leading to debates about the fungal ancestral state[19,24–26]. Taken together, our results provide evidence that conversion from homothallism to heterothallism is possible, and suggest that the predominance of a particular sexual strategy may vary within different clades.

Although the finding of MAT genes in supposedly asexual fungi has been previously reported[15,27–32] and genes related to sexual reproduction have been found in the 'asexual' yeast *Candida albicans*, this report is the first comprehensive survey of sexual reproduction genes in two different filamentous fungi thought to be asexual. In addition, our results provide an experimentally supported evolutionary model associating large-scale synteny and genome rearrangement with a specific and significant difference in biology between these aspergilli. These results for *A. fumigatus* and *A. oryzae* have important and specific potential implications for health and industry. The lack of a sexual cycle in *A. fumigatus* and *A. oryzae* has precluded classical genetic analysis, impeding efforts to study these organisms and necessitating the use of the relatively distant *A. nidulans* as a genetic model. The possibility for mating—still speculative at this stage—raises the medically and industrially important potential for developing genetic tools for these fungi.

Conserved non-coding sequences

Detecting and characterizing conservation of sequences outside of protein coding regions is a promising method for identifying potential functional elements. Regulation in yeast has been extensively studied; however, in the aspergilli few transcription factor binding sites have been experimentally verified. Comparing the three aspergilli provides an opportunity to identify the most constrained functional elements.

To do so we aligned three-way orthologous genes including 1 kb of sequence upstream and downstream using Mlagan[33]. Strict filters were then applied to delineate unambiguous orthologous intergenic regions (see Methods). Given the divergence of the aspergilli, it is expected that intergenic regions would not show significant conservation, and frequently this was found to be the case. However, in many instances, blocks of nearly perfect three-way conservation

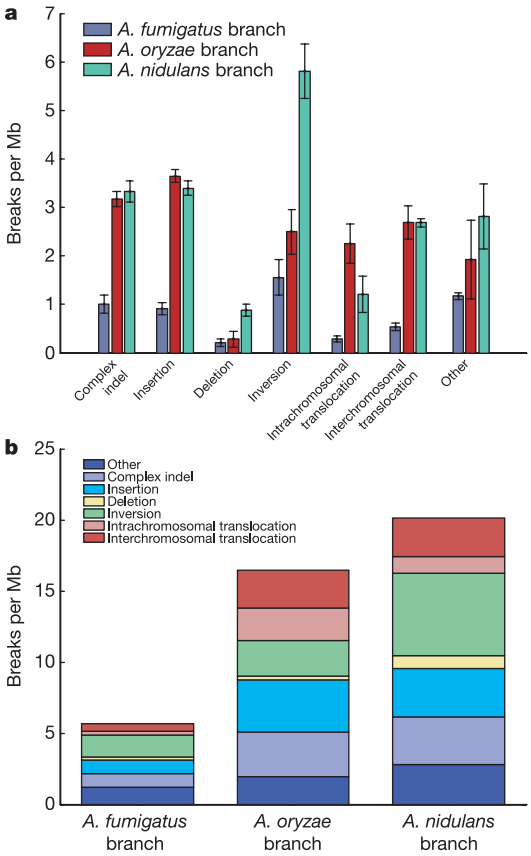

Figure 3 | Rates of branch-specific rearrangements. a, The rates of different breaks broken down by break type for each branch. Bars represent minimum and maximum values obtained using either of the two non-target genomes as reference (see Methods). **b**, A stacked plot of the same data showing the relative contribution of break types within each branch for all three branches. See text and Methods for more details.

© 2005 Nature Publishing Group

are observed. An example for one intergenic region is shown in Fig. 5a.

To assess which regions were conserved owing to purifying selection rather than neutral mutation or chance, we developed models for alignments of neutral and random sequence. Unlike mammals, where ancient conserved repeats provide a natural model for neutral evolution, few such repeats exist in the aspergilli (see below). Instead, we synthesized alignments of neutral sequence by concatenating

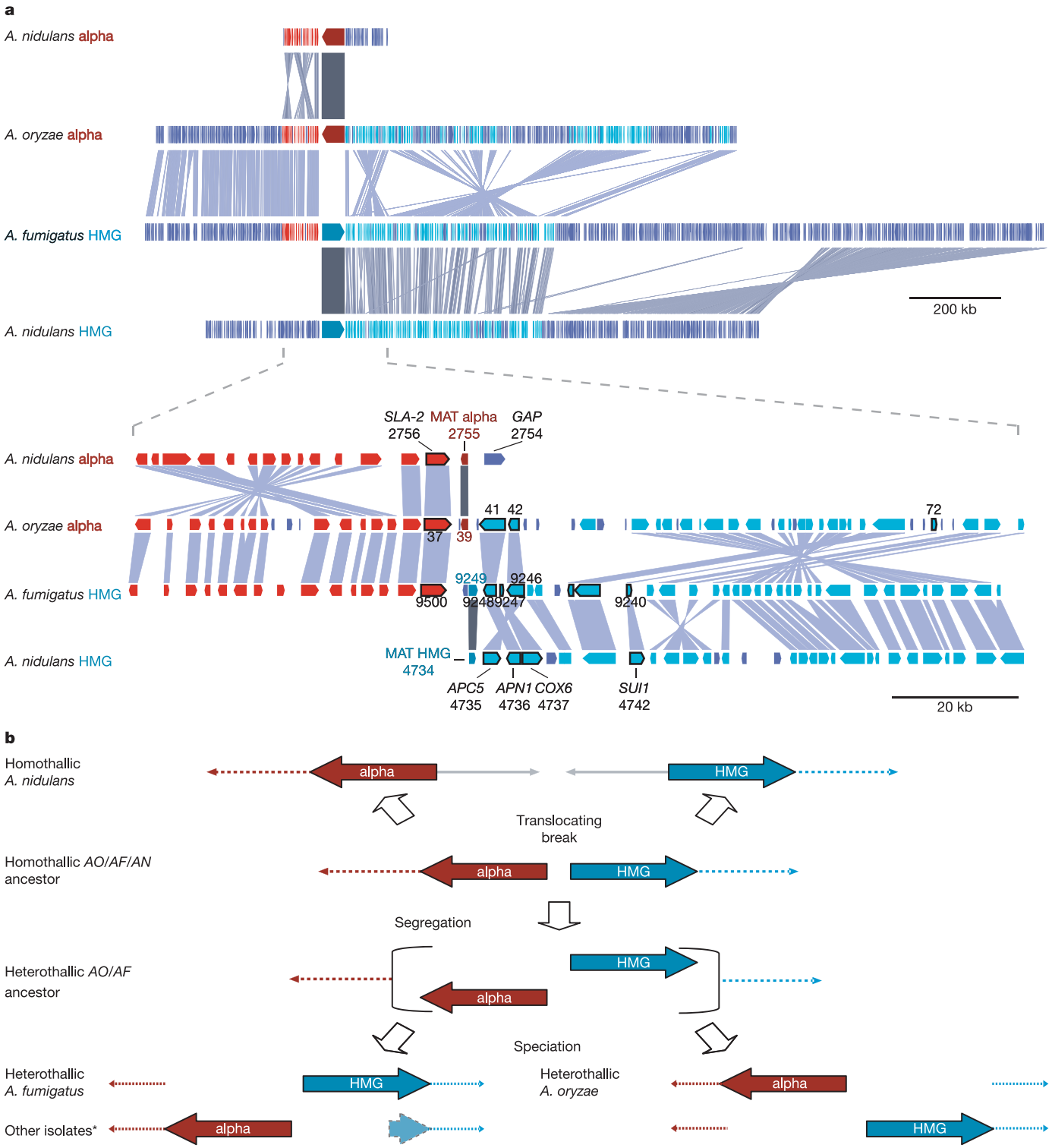

Figure 4 | Comparison and evolutionary model of *Aspergillus* MAT loci. **a**, Conserved synteny between loci. Grey lines indicate predicted orthologues. Red genes indicate orthologues from the left flank (as drawn) of the *A. nidulans* alpha locus with the left flanks of the *A. fumigatus* and *A. oryzae* loci. Cyan genes indicate orthologues with the right flank of the *A. nidulans* HMG locus. The bottom panel shows the region near mating-type genes. Genes labelled and outlined in black are associated with MAT loci in other fungi. Only partial accession numbers (suffixes) are shown in the figure. For full accession numbers, the numbers shown in the panel should replace the asterisks in the following examples: *A. nidulans* (AN****.1); *A. fumigatus* (59.m0****); *A. oryzae* (AO0703270000**). **b**, Model of structural evolution of the MAT loci. Braces represent multiple haplotypes at the same genomic locus. The experimental identification of other isolates (indicated by an asterisk) was reported in another study[15]. The light blue arrow indicates a 360-bp HMG gene fragment. *AF*, *A. fumigatus*; *AN*, *A. nidulans*; *AO*, *A. oryzae*.

© 2005 Nature Publishing Group

randomly selected aligned columns of fourfold degenerate sites. We also controlled for chance alignment introduced by potential aligner bias by aligning randomly selected intergenic regions. Using a simple conservation scoring function we calculated maximal scoring subsequences and compared results between orthologous regions and the control models (Fig. 5b). A noteworthy aspect of the data in Fig. 5b is the similarity between the neutral and random models. According to our models, neutral sequence is effectively saturated for mutation, confirmed by an independent analysis of synonymous sites in protein coding sequences.

Comparing results between real and control alignments, we selected a minimum score of 22 for regions unlikely ($P < 0.015$) to be conserved by neutral evolution or chance. We denote a subsequence scoring above this threshold as a high-scoring conserved sequence (HCS). On the basis of this analysis, we predict 5,801 HCSs corresponding to ~2% of alignable orthologous intergenic regions.

Prediction of functional motifs

We expect HCSs to be enriched for functional elements. The challenge is to discover these functional elements and make testable predictions about their biological functions. In preliminary analyses, several conserved regions could be identified as known functional elements. For example, we observed conservation delimiting a known 3′ untranslated region (UTR) element of the *A. nidulans areA* gene that regulates messenger RNA stability in response to cellular nitrogen levels[34]. We also identified three TPP binding riboswitches, one of which has not been described in *Aspergillus* (Supplementary Information).

To enrich computationally HCSs for sequences corresponding to functional elements, and to derive clues about their biological functions, we modified the approach used by ref. 35 (see Methods). Briefly, we identified common subsequences (or 'patterns') that appeared in at least four HCSs across all three *Aspergillus* genomes. These patterns were searched in three-way conserved orthologues to identify genes in which the subsequence occurred in the 500-bp upstream or downstream regions (a 'co-occurrence'). A number of conservation criteria were then applied (see Methods). We identified a total of 69 conserved patterns ('conpats'), occurring in at least four HCSs, that showed enrichment for co-occurrences and exhibited a bias for occurring 500 bp upstream or downstream of genes.

The results of this analysis for the 35 most common patterns are shown in Fig. 6 (all 69 patterns available in Supplementary Information). These include motifs that match known or predicted *Aspergillus* or other fungal functional sequences. For example, CPCA/GCN4, the master regulator of the cross pathway control system in fungi, is known to bind to the palindromic site TGASTCA[36]. In yeast, microarray studies have identified 539 genes probably regulated by GCN4 that show a preference for amino acid biosynthetic genes and several ribosomal proteins and translation factors[37]. One of the patterns identified by our analysis (ID 2483) matched the CPCA binding site, co-occurred preferentially upstream of genes, and was enriched in genes associated with amino acid transport and metabolism (COG category E), and translation, ribosomal structure and biogenesis (category J). Furthermore, the 19 genes with co-occurrences of this pattern include 7 (37%) predicted orthologues to the 539 known yeast regulated genes,

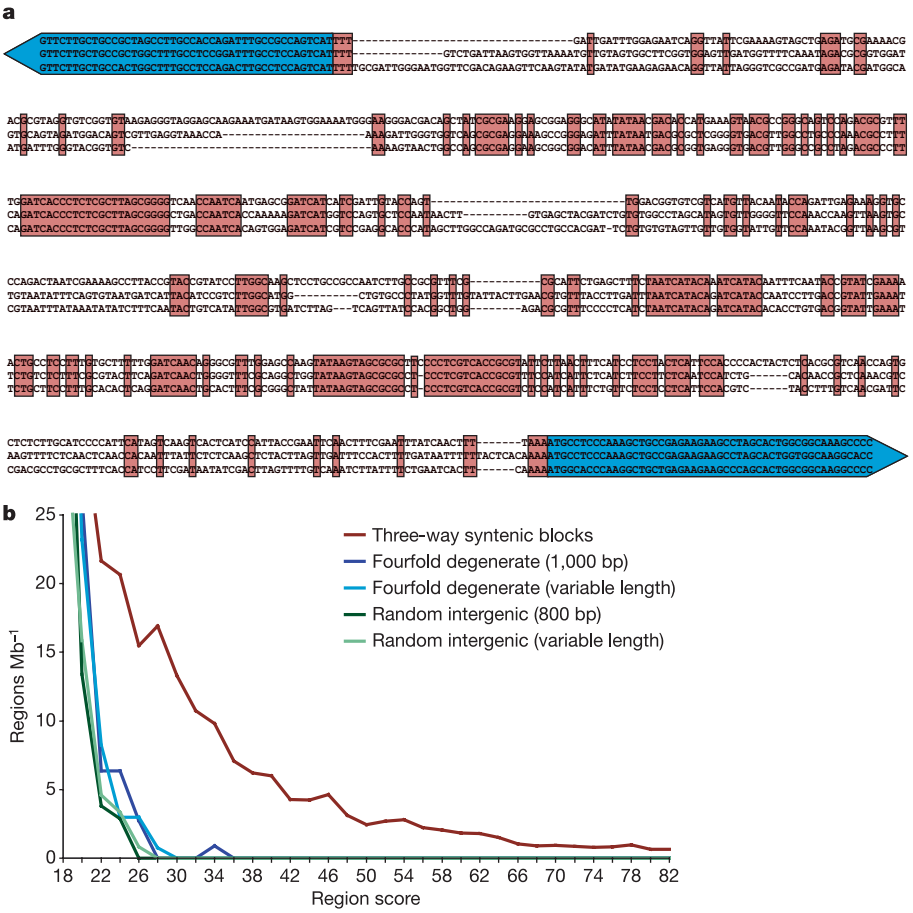

Figure 5 | Active conservation of non-coding regions. a, Example region between a conserved pair of orthologous histone H2A and H2B genes (left and right blue arrows). The three lines from top to bottom correspond to the sequences of *A. nidulans*, *A. fumigatus* and *A. oryzae* aligned using Mlagan. Letters on the red background indicate 100% conserved bases. **b**, Conservation scores of maximal subsequences for observed intergenic alignments (red), and models of neutral and random sequence alignment (both fixed and variable length).

© 2005 Nature Publishing Group

representing a sixfold enrichment (P-value $< 1 \times 10^{-5}$). This includes two known *A. nidulans* CPCA regulated genes (*trpB*[38] and *hisHF*[39]). This pattern probably corresponds to the known *Aspergillus* CPCA binding site.

A second pattern shows strong correspondence with the binding site for Puf family genes[40]. Puf proteins regulate mRNA translation and mRNA decay through interactions with 3′ UTR sequences. In *Saccharomyces cerevisiae*, which has five Puf genes, Puf3p has been shown to bind specifically mRNAs encoding mitochondrial proteins[40] and requires a 3′ UTR motif with consensus UGUANAUA[40]. Four different patterns identified by our analysis (ID 1710, 2077, 1144 and 2378; see the full table in Supplementary Information) match or include the Puf binding motif and display a strong downstream bias. Three also show enrichment for predicted orthologues in *S. cerevisiae* that localize to mitochondria. Taking all four patterns together, we find a 6.8-fold enrichment ($P < 4.3 \times 10^{-11}$) for genes with orthologues to yeast mitochondrial genes. In addition, we find a threefold enrichment ($P < 0.0003$) for genes with yeast orthologues predicted to be bound by Puf genes in a genome-wide affinity tag analysis[40]. Although a functional role for the Puf family has not been experimentally demonstrated in *Aspergillus*, all three genomes possess five loci with 5–8 Puf domains, as predicted by HMMER and PFAM (including one with a predicted RNA binding domain, as with Puf1p and Puf2p in yeast). Together these data suggest that, as in yeast, Puf genes may bind to and regulate mitochondrial mRNAs in the aspergilli.

Only a small number of transcription factor binding sites and control elements are known for filamentous fungi in general, including *Aspergillus*. These predicted patterns are thus promising targets for future experimental validation.

Regulatory upstream open reading frames

A significant proportion (32%) of HCSs are conservatively predicted to lie within transcribed but untranslated regions of genes (UTRs), consistent with the known role of UTRs in regulating gene expression, particularly mRNA translation (for example, Puf binding domains). One important class of translational control elements is short upstream open reading frames (uORFs) in 5′ UTRs[41], which can regulate the expression of downstream protein-coding genes in several ways. First, they can modulate the efficiency of ribosome

ID	Conpat consensus	Similar sequence	No. 5' sites	No. 3' sites	5' /3' bias?	Aligned (%)	Strand bias?	Enriched in COG categories (no. genes in category)
1821		bHLH	115	8	up***	35	Yes***	J: Translation, ribosomal structure, biogenesis (16)*
1622		*areA/gln3*	25	3	up*	60		
234			46	11	up**	35		A: RNA processing, modification (8)
1934		bHLH	36	3	up***	47	Yes*	I: Lipid transport, metabolism (10)**
1710		Puf	5	20	down*	30	Yes***	
896			14	2	up	43	Yes	
1602			34	0	up***	50		
1879			24	1	up**	63	Yes	
1132			14	2	up	36	Yes***	
2683			27	1	up***	48	Yes*	
2223			5	16	down	63	Yes***	
1291		yAP1	21	1	up**	62	Yes	
992			1	21	down***	86	Yes***	
2611			0	17	down***	59		
1739			9	0	up	44	Yes*	J: Translation, ribosomal structure, biogenesis (4)
418			33	1	up***	45		
2077		Puf	1	9	down	56	Yes	
973			74	6	up***	36	Yes*	A: RNA processing, modification (12)*
2488			37	4	up***	38	Yes**	T: Signal transduction mechanisms (11)**
1144		Puf	2	11	down	36	Yes***	
1635			13	1	up	69	Yes***	J: Translation, ribosomal structure, biogenesis (8)***
1940			2	19	down**	47	Yes***	D: Cell cycle/division, chromosomal partitioning (5)*
2199		bHLH	40	3	up***	30		
1426			14	1	up	50		C: Energy production, conversion (5)*
2525		*stuA/mbp1*	18	1	up*	39	Yes	L: Replication, recombination, repair (10)***
813		yATF-B	11	1	up	45		
1650			11	0	up	73		J: Translation, ribosomal structure, biogenesis (4)
2093			44	5	up***	30	Yes***	I: Lipid transport, metabolism (8)
1998			25	1	up***	40	Yes	I: Lipid transport, metabolism (7)*
2497			0	8	down	100	Yes**	
751			25	2	up**	40		
2483		*cpcA/gcn4*	19	4	up	26		E: Amino acid transport, metabolism (7)* J: Translation, ribosomal structure, biogenesis (7)**
2249			36	0	up***	25		A: RNA processing, modification (7) J: Translation, ribosomal structure, biogenesis (13)***
1059			12	1	up	92		
2451			0	10	down*	80		

Figure 6 | Selected conserved patterns. Column one shows the conpat unique ID. Column two shows the sequence logo representation of conpat weight matrix. Column three shows fungal binding factors with sequence similarity to the conpat. Columns four and five show the number of genes with a co-occurrence of conpat upstream and downstream. Column six shows the preference for co-occurring preferentially 5′ or 3′ of the gene. Column seven shows the fraction of co-occurrences overlapping three-way conserved regions. Column eight shows the preference for co-occurring on a particular strand relative to the gene. Column nine shows COG categories showing significant enrichment (the number of genes with co-occurring conpats in the category is indicated in parentheses). Enrichment results for yeast orthologue cellular location are available in the Supplementary Information. Single asterisk, $P < 1 \times 10^{-3}$; double asterisk, $P < 1 \times 10^{-4}$; triple asterisk, $P < 1 \times 10^{-5}$. bHLH, basic helix-loop-helix.

© 2005 Nature Publishing Group

re-initiation at downstream start codons in a manner dependent on cellular state. Second, uORFs can produce *cis*-acting peptides that stall ribosomes. Finally, the presence of uORFs can affect mRNA stability. Functional uORFs can be as short as three amino acids and occur at varying distances and multiplicity upstream of the protein-coding gene, occasionally overlapping the downstream start codon. Functional uORFs have been reported in a range of species including plants, animals and fungi[41]. In *Aspergillus*, a small number of genes with validated uORFs have been reported (Supplementary Information). To determine the full extent with which they may regulate gene expression, we analysed the three *Aspergillus* genomes for uORFs.

We identified UTR sequences using expressed sequence tag (EST) alignments for *A. nidulans* genes, and examined them for open reading frames (see Methods). Of 1,606 genes with identified UTRs, 21% (358) have upstream ORFs. A similar proportion was found (18% or 82 out of 463) when we restricted our analysis to three-way orthologues for which the start codons align exactly within multiple alignments, suggesting that these uORFs are not due to mis-annotation. A corresponding analysis of genes with ESTs in the *N. crassa*, *F. graminearum* and *Magnaporthe grisea* genomes found uORFs associated with 22%, 10% and 16% of genes, respectively. We further extended the analysis in *A. nidulans* using a conservative estimate of 5′ UTR length (see Methods), and identified an additional 958 genes with potential uORFs of which 165 genes have three-way orthologues. In total, 1,316 genes in *A. nidulans* are predicted to possess uORFs.

Not all identified uORFs have a detectable impact on gene expression[41]. To enrich the set of predicted uORFs for those likely to be functional, we looked for those conserved in all three aspergilli. On the basis of a strict criterion requiring alignment of the uORF start and stop codons, we find 38 conserved uORFs (13% of 331 genes with uORFs and predicted orthologues) (Supplementary Information). Of these corresponding *Aspergillus* genes, 14 have predicted uORFs upstream of orthologues in *N. crassa*, *F. graminearum* or *M. grisea*. Additionally, three also have predicted orthologues in *S. cerevisiae* with uORFs conserved across four related yeast species.

These 38 conserved uORFs represent strong candidates for experimental investigation. As a preliminary validation, we tested two novel conserved uORFs for their ability to modulate protein synthesis *in vitro* (Fig. 7). Briefly, oligonucleotides containing each uORF were fused to a luciferase reporter gene, and controls were constructed with disabled uORF and/or reporter gene start codons. Differential expression in a cell-free translation system between intact and control constructs measures the impact of the uORF on translation. This system can detect small (twofold) changes in translation, and can discriminate uORFs that do not reduce translation *in vivo* from those that do (see Methods). As can be seen in Fig. 7, both uORFs tested display a 5–10-fold repressive effect on the translation of the downstream reporter gene.

These results provide the first genome-wide list of predicted conserved uORFs for any organism, and suggest that uORFs could have a substantial role in regulating gene expression in *Aspergillus*. Previous reports estimate that 2–4% of genes in *S. cerevisiae* contain uORFs[42], whereas a review of sequences in UTRdb predicted that 5–10% of eukaryotic UTRs contain ORFs[43]. Our results predict that in filamentous fungi the proportion may be twice as high.

Aspergillus physiology

Peroxisomes are organelles containing enzymes for the breakdown of fatty acids (β-oxidation), removal of hydrogen peroxide and synthesis of cholesterol and bile acids. Peroxisomes have critical roles in fungi where they are involved in growth, secondary metabolism and pathogenesis. In mammals, defects can lead to developmental and neurological disorders. Proteins are targeted to the peroxisome either by a carboxy-terminal tripeptide sequence or an amino-terminal nine-amino-acid sequence. Using these signals we predicted peroxisomal proteins in *Aspergillus* (Supplementary Information). Our analysis reveals peroxisomes in *Aspergillus* to be more similar to mammals than yeasts in two respects. First, our data suggest that the aspergilli, like mammalian cells, perform β-oxidation in both peroxisomes and mitochondria and possess two sets of genes for all β-oxidation enzymes targeted to both the mitochondria and the peroxisome, as supported by recent experimental results[44]. In contrast, *S. cerevisiae* metabolizes fatty acids fully to acetyl-CoA only in peroxisomes[45]. Second, *Aspergillus* peroxisomes are more similar to those of mammals than those of yeasts in that they possess putative peroxisomal acyl-CoA dehydrogenases. In addition, all three aspergilli appear to encode both mitochondrial and peroxisome forms of an ATP-dependent protease of the LON (La domain) family associated with peroxisomes in mammals[46]. *S. cerevisiae* has a single copy of this protease (*Pim1*/*Lon1*) targeted to mitochondria[47].

One of the hallmarks of the filamentous fungi is their ability to undergo polarized hyphal growth. This requires positional cues that mark polarized growth sites, locally activating Rho-related GTPase signalling modules that promote cytoskeletal reorganization[48]. The three aspergilli possess the expected genes involved in signalling and cytoskeletal organization for polarized growth, but there is a marked lack of known positional markers (such as the yeast bud site markers Bud3p, Bud8p and Bud9p; see Supplementary Information). Proteins implicated in the transport or modification of bud markers, including Axl1p, Rax1p and Bud7p, were predicted, however. This suggests that filamentous fungi mark polarized growth sites with positional cues, but that the markers themselves may consist of novel cell wall proteins.

Most of the interspersed repeats in all three genome sequences correspond to relics of transposable elements (see Supplementary Information). Surprisingly, only 1.3% of the largest genome

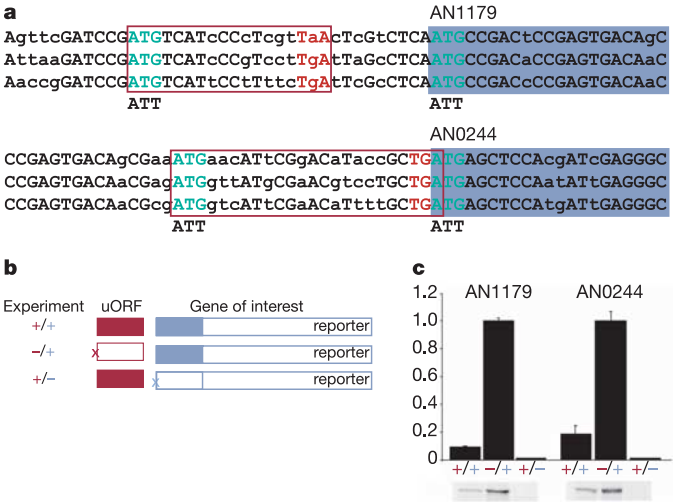

Figure 7 | Prediction and validation of conserved uORFs. Conserved uORFs show a 5–10 times repressive effect on reporter gene translation. **a**, Alignments of two tested uORFs. uORFs are shown in purple boxes, and protein-coding genes in a blue background. Conserved bases are in upper case and start/stop codons are highlighted. **b**, Experimental design. The *A. nidulans* sequence from 26 nucleotides upstream of each uORF to four codons in the protein-coding gene was fused with a firefly luciferase gene. Controls were generated with start codons for both the uORF and the luciferase gene (+/+), the luciferase gene only (−/+), and the uORF only (+/−). Starts were deleted by alteration to ATT. mRNA from each construct was used to programme a cell-free translation system. **c**, Results of translation assays. The luciferase activity of all constructs (normalized to the −/+ construct) is shown on the *y* axis. Error bars show the average of the absolute deviation from the mean. The autoradiogram shows ^{35}S-Met-labelled firefly luciferase obtained by *in vitro* translation of the same mRNAs.

© 2005 Nature Publishing Group

assembly of *A. oryzae* consists of transposable elements, compared to 3% of *A. nidulans* and *A. fumigatus*. All three genomes contain essentially all major classes of eukaryotic transposable elements, although the overall variety is relatively low. A number of unusual features were also observed in *Aspergillus* transposable elements (Supplementary Information). Copy numbers per family range from 1 to ~100; there are also many fragments and 'footprints' (evidence of recombinational excision). In addition, some members of all transposable element families longer than ~400 nucleotides are characterized by numerous C to T transitions. In *A. nidulans* and *A. fumigatus*, these predominantly affect cytosines in CpG and CpA doublets (no preference is apparent in *A. oryzae*). Moreover, repeat density is correlated with A + T richness in all three species (Fig. 2). The predominance of transition mutations is consistent with the operation of RIP (repeat-induced point mutation)[49], and all three aspergilli have a single predicted homologue (called DmtA) to the DNA methyltransferase *rid* that is essential for RIP in *N. crassa*[49]. Apart from the putative *rid* homologue, no additional DNA methyltransferase genes were identified, consistent with failures to demonstrate methylation in these fungi. Although RIP has not been demonstrated in any *Aspergillus* species, if active it may be more similar to the mild form in *M. grisea*[49], as many transposable elements in these species are mutation-free.

Conclusion and perspective

The *A. nidulans* genome sequence and our comparative analysis with the genome sequences of *A. fumigatus* and *A. oryzae* have shed new light on the physiology of these fungi, as well as insight into aspects of genome evolution and gene regulation likely to be common to all eukaryotes. These results represent the initial step in realizing the full potential of these genomes. As a result of the genome analysis, efforts are underway to cross different isolates of *A. fumigatus* and *A. oryzae*. The identified conserved sequences also represent a rich set of targets for further experimental investigation. These efforts and ongoing sequencing projects for additional aspergilli promise to change fundamentally our understanding of this important group of medically, industrially and scientifically relevant fungi.

METHODS

Complete details of the methods used are available in Supplementary Information.

***A. nidulans* sequencing, assembly and analysis.** The *A. nidulans* genome strain FGSC A4 was sequenced by the WGS method to a depth of 10×. An additional 3× sequence coverage was provided by Monsanto (http://www.monsanto.com/). All sequence was assembled using Arachne. The *A. nidulans* genome was annotated as described in Supplementary Information. *A. fumigatus* and *A. oryzae* were assembled and annotated as described separately[2,3].

Phylogenetic analysis. A total of 3,034 predicted orthologues among *A. nidulans*, *A. oryzae*, *A. fumigatus*, *N. crassa* and *F. graminearum* were aligned at the protein level, back translated to DNA codons, and large gaps (>9 bp) were removed. Random sets of 20 DNA alignments were concatenated and passed to Phylip to generate 100 bootstrap replicates and a consensus maximum parsimony tree. Maximum likelihood trees were calculated on each replicate and a consensus tree was produced. Repeating with 1,000 bootstrap samples led to essentially identical results. For rooting with *C. immitis*, *C. immitis* orthologous CDS regions based on TBlastN were retrieved, translated and aligned at the protein level with the aligned portions of the *Aspergillus* genes. Maximum parsimony trees were then generated and filtered for those with 100% bootstrap values at all nodes. The *C. immitis* sequence is available at http://www.broad.mit.edu/annotation/fungi/coccidioides_immitis/.

Hierarchical synteny mapping and branch-specific rearrangements. Protein homology anchors were detected using BlastP and filtered to retain only hits scoring >80% of the score of the best hit to each query protein. Contiguous sets of homologous proteins with conserved order and orientation were grouped into clusters. Pairs of clusters were then merged into successively larger clusters by tolerating successively larger breaks between clusters. Branch-specific breaks were determined by identifying regions without breaks between a reference species and query species, and then identifying breaks in that region between the reference and the third (target) species. Such breaks were considered specific to the third species. Breaks were classified according to the pattern of apparent rearrangement.

Identification of non-coding conserved sequences. Genomic sequence for predicted three-way orthologues, including 1 kb upstream and downstream, were multiply aligned using Mlagan[33]. An additive scoring function was used to identify maximal scoring subsequences. A cutoff of 22 was used to define HCSs unlikely to occur by chance ($P < 0.015$) according to models of neutral sequence and random sequence alignment. To model alignments of neutrally evolving sequence, aligned columns of fourfold degenerate sites were selected randomly and concatenated. To model alignments of random sequence, randomly selected intergenic regions from each genome were aligned. Fixed and variable length alignments were generated for both models. For each model, 1,000 simulated alignments were generated and maximal scoring subsequences were identified. The number of subsequences for each score was normalized by the number of aligned nucleotides. These rates were used to determine the score cutoff above.

Prediction of functional motifs. Each HCS was represented as a position-specific probability matrix (PSPM) derived from the three-way alignment. Each PSPM was compared to each other PSPM and matching PSPMs were clustered. Local multiple alignments for each cluster were generated and the resulting multiple alignments and corresponding weight matrices were termed conpats. For each conpat, we used the corresponding weight matrix with MAST to identify instances where the conpat co-occurred upstream or downstream of orthologous genes in all three of the aspergilli. A series of conservation tests was then applied to the set of predicted co-occurrences for each conpat as described in Supplementary Information.

Prediction and validation of upstream open reading frames. Genome sequences for three-way orthologues, including 1,000 bp upstream and downstream, were multiply aligned using Mlagan. For 25% of *A. nidulans* genes 5′ UTR sequences were predicted from EST alignments. On the basis of the length distribution of these EST-predicted UTRs, we used 60 bp upstream of predicted AUG codons as a conservative estimate of 5′ UTRs for genes lacking ESTs. When no ESTs were available, UTRs for orthologues where considered when all three annotated start codons aligned within 40 bp. We identified uORFs ≥12 bp, with a maximum 1-bp overlap with the protein-coding gene's ATG. Conserved uORFs were identified as those for which the start and stop codons were exactly aligned within the multiple alignments. To experimentally validate uORFs, synthetic oligonucleotides containing each uORF, 26 nucleotides upstream of the uORF, the region between the uORF and the protein coding AUG, and the first four codons of the protein coding gene were fused in frame with a firefly luciferase gene. Three different control constructs were also generated. Capped and polyadenylated synthetic mRNA were prepared from each construct and equal amounts were used to programme cell-free extracts from *N. crassa*. Differential translation between the intact construct and the controls was measured using a luciferase activity assay as well as through the production of ^{35}S-Met-labelled firefly luciferase obtained by *in vitro* translation of the same mRNAs.

The *A. nidulans* genome sequence is available at http://www.broad.mit.edu/ and has been deposited at DDBJ/EMBL/GenBank under the project accession AACD00000000.

Received 31 May; accepted 19 October 2005.

1. Timberlake, W. E. & Marshall, M. A. Genetic engineering of filamentous fungi. *Science* **244**, 1313–1317 (1989).
2. Nierman, W. *et al.* Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* doi:10.1038/nature04332 (this issue).
3. Machida, M. *et al.* Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* doi:10.1038/nature04300 (this issue).
4. Pontecorvo, G. & Kafer, E. Genetic analysis based on mitotic recombination. *Adv. Genet.* **9**, 71–104 (1958).
5. Pontecorvo, G., Roper, J. A., Hemmons, L. M., Macdonald, K. D. & Bufton, A. W. The genetics of *Aspergillus nidulans*. *Adv. Genet.* **5**, 141–238 (1953).
6. Peterson, S. W. in *Integration of Modern Taxonomic Methods for* Penicillium *and* Aspergillus *Classification* (eds Samson, R. A. & Pitt, J. I.) 323–355 (Harwood Academic Publishers, Amsterdam, The Netherlands, 2000).
7. Rokas, A., Williams, B. L., King, N. & Carroll, S. B. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804 (2003).
8. Verweij, P. E. *et al.* Phylogenetic relationships of five species of *Aspergillus* and related taxa as deduced by comparison of sequences of small subunit ribosomal RNA. *J. Med. Vet. Mycol.* **33**, 185–190 (1995).
9. Dujon, B. *et al.* Genome evolution in yeasts. *Nature* **430**, 35–44 (2004).
10. Eichler, E. E. & Sankoff, D. Structural dynamics of eukaryotic chromosome evolution. *Science* **301**, 793–797 (2003).
11. Coghlan, A. & Wolfe, K. H. Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. *Genome Res.* **12**, 857–867 (2002).

© 2005 Nature Publishing Group

12. Burt, D. W. *et al.* The dynamics of chromosome evolution in birds and mammals. *Nature* **402,** 411–413 (1999).
13. Sharakhov, I. V. *et al.* Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus*. *Science* **298,** 182–185 (2002).
14. Slamovits, C. H., Fast, N. M., Law, J. S. & Keeling, P. J. Genome compaction and stability in microsporidian intracellular parasites. *Curr. Biol.* **14,** 891–896 (2004).
15. Paoletti, M. *et al.* Evidence for sexuality in the opportunistic fungal pathogen *Aspergillus fumigatus*. *Curr. Biol.* **15,** 1242–1248 (2005).
16. Metzenberg, R. L. & Glass, N. L. Mating type and mating strategies in *Neurospora*. *Bioessays* **12,** 53–59 (1990).
17. Dyer, P. S. in *Proc. 7th Int. Mycol. Congr.* Abstr. 1095, 331 (International Mycological Congress, Oslo, 2002).
18. Dyer, P. S., Paoletti, M. & Archer, D. B. Genomics reveals sexual secrets of *Aspergillus*. *Microbiology* **149,** 2301–2303 (2003).
19. Yun, S. H., Berbee, M. L., Yoder, O. C. & Turgeon, B. G. Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. *Proc. Natl Acad. Sci. USA* **96,** 5592–5597 (1999).
20. Poggeler, S. Genomic evidence for mating abilities in the asexual pathogen *Aspergillus fumigatus*. *Curr. Genet.* **42,** 153–160 (2002).
21. Butler, G. *et al.* Evolution of the MAT locus and its Ho endonuclease in yeast species. *Proc. Natl Acad. Sci. USA* **101,** 1632–1637 (2004).
22. Geiser, D. M., Timberlake, W. E. & Arnold, M. L. Loss of meiosis in *Aspergillus*. *Mol. Biol. Evol.* **13,** 809–817 (1996).
23. Geiser, D. M., Frisvad, J. & Taylor, J. Evolutionary relationships in *Aspergillus* section *Fumigati* inferred from partial β-tubulin and hydrophobin DNA sequences. *Mycologia* **90,** 831–845 (1998).
24. Perkins, D. D. Mating-type switching in filamentous ascomycetes. *Genetics* **115,** 215–216 (1987).
25. Nauta, M. J. & Hoekstra, R. F. Evolution of reproductive systems in filamentous ascomycetes. II. Evolution of hermaphroditism and other reproductive strategies. *Heredity* **68,** 537–546 (1992).
26. Coppin, E., Debuchy, R., Arnaise, S. & Picard, M. Mating types and sexual development in filamentous ascomycetes. *Microbiol. Mol. Biol. Rev.* **61,** 411–428 (1997).
27. Lucas, J. A., Dyer, P. S. & Murray, T. D. Pathogenicity, host specificity, and population biology of *Tapesia* spp, causal agents of eyespot disease of cereals. *Adv. Bot. Res.* **33,** 225–258 (2000).
28. Kang, S., Chumley, F. G. & Valent, B. Isolation of the mating-type genes of the phytopathogenic fungus *Magnaporthe grisea* using genomic subtraction. *Genetics* **138,** 289–296 (1994).
29. Kerenyi, Z., Moretti, A., Waalwijk, C., Olah, B. & Hornok, L. Mating type sequences in asexually reproducing fusarium species. *Appl. Environ. Microbiol.* **70,** 4419–4423 (2004).
30. Magee, B. B. & Magee, P. T. Induction of mating in *Candida albicans* by construction of MTLa and MTLα strains. *Science* **289,** 310–313 (2000).
31. Hull, C. M., Raisner, R. M. & Johnson, A. D. Evidence for mating of the "asexual" yeast *Candida albicans* in a mammalian host. *Science* **289,** 307–310 (2000).
32. Sharon, A. *et al.* An asexual fungus has the potential for sexual development. *Mol. Gen. Genet.* **251,** 60–68 (1996).
33. Brudno, M. *et al.* LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13,** 721–731 (2003).
34. Morozov, I. Y., Martinez, M. G., Jones, M. G. & Caddick, M. X. A defined sequence within the 3′ UTR of the *areA* transcript is sufficient to mediate nitrogen metabolite signalling via accelerated deadenylation. *Mol. Microbiol.* **37,** 1248–1257 (2000).
35. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423,** 241–254 (2003).
36. Hoffmann, B., Valerius, O., Andermann, M. & Braus, G. H. Transcriptional autoregulation and inhibition of mRNA translation of amino acid regulator gene *cpcA* of filamentous fungus *Aspergillus nidulans*. *Mol. Biol. Cell* **12,** 2846–2857 (2001).
37. Natarajan, K. *et al.* Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Mol. Cell. Biol.* **21,** 4347–4368 (2001).
38. Eckert, S. E., Kubler, E., Hoffmann, B. & Braus, G. H. The tryptophan synthase-encoding *trpB* gene of *Aspergillus nidulans* is regulated by the cross-pathway control system. *Mol. Gen. Genet.* **263,** 867–876 (2000).
39. Valerius, O. *et al.* Regulation of *hisHF* transcription of *Aspergillus nidulans* by adenine and amino acid limitation. *Fungal Genet. Biol.* **32,** 21–31 (2001).
40. Gerber, A. P., Herschlag, D. & Brown, P. O. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol.* **2,** E79 (2004).
41. Geballe, A. P. & Sachs, M. S. in *Translational Control of Gene Expression* (eds Sonenberg, N., Hershey, J. W. B. & Mathews, M. B.) 595–614 (Cold Spring Harbor Laboratory Press, New York, 2000).
42. McCarthy, J. E. Posttranscriptional control of gene expression in yeast. *Microbiol. Mol. Biol. Rev.* **62,** 1492–1553 (1998).
43. Mignone, F., Gissi, C., Liuni, S. & Pesole, G. Untranslated regions of mRNAs. *Genome Biol.* **3,** REVIEWS0004 (2002).
44. Maggio-Hall, L. A. & Keller, N. P. Mitochondrial β-oxidation in *Aspergillus nidulans*. *Mol. Microbiol.* **54,** 1173–1185 (2004).
45. Hiltunen, J. K. *et al.* The biochemistry of peroxisomal β-oxidation in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* **27,** 35–64 (2003).
46. Kikuchi, M. *et al.* Proteomic analysis of rat liver peroxisome: presence of peroxisome-specific isozyme of Lon protease. *J. Biol. Chem.* **279,** 421–428 (2004).
47. Balzi, E., Wang, M., Leterme, S., Van Dyck, L. & Goffeau, A. PDR5, a novel yeast multidrug resistance conferring transporter controlled by the transcription regulator PDR1. *J. Biol. Chem.* **269,** 2206–2214 (1994).
48. Harris, S. D. & Momany, M. Polarity in filamentous fungi: moving beyond the yeast paradigm. *Fungal Genet. Biol.* **41,** 391–400 (2004).
49. Galagan, J. E. & Selker, E. U. RIP: the evolutionary cost of genome defense. *Trends Genet.* **20,** 417–423 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors would like to thank M. Zody, X. Xie, M. Kellis, M. Kamal, J. Taylor, G. Turgeon and E. Lander for numerous helpful conversations, comments and critical readings of the paper. We also thank all members of the sequencing platform at the Broad Institute. We thank R. Dean for providing the BAC library used in sequencing. We especially thank M. Brudno for his help in using Mlagan to align the three genomes, and R. Morris for his many contributions to the *A. nidulans* sequencing project and subsequent analyses. This work was supported in part by grants from the NIH, as well as the NIH Research Supplement for Underrepresented Minorities, the DFG, the CMPB and the BBSRC.

Author Contributions B.W.B. developed and led the *A. nidulans* sequencing project. J.B. performed the assembly of *A. nidulans*. C.C. and J.E.G. analysed the phylogenic relationship of the three organisms. J.E.G. performed the analysis of conserved synteny and genome evolution. R.F., B.M. and P.D. performed the comparative analysis of sexual reproduction genes. J.E.G., M.P., P.D. and B.M. analysed the mating-type loci. S.E.C. and J.E.G. analysed conserved non-coding sequences and computationally analysed uORFs. S. Batzoglou and S.-I.L. assisted with the use of Mlagan. M.B., C.C.S. and M.S.S. performed the experimental uORF validation. M.C., M.H., G.H.B., O.D. and C.D. analysed transcription, known transcription factors and binding sites. A.P. and S.G.J. predicted non-coding RNAs using PFAM and RFAM. J.C., V.K., J.J., S.P. and J.E.G. analysed repeat sequences and RIP. M. Farman analysed telomeres and subtelomeric gene content. M.H. analysed the Aspergillus peroxisomes. S.H. and M. Momany analysed hyphal growth and RhoGTPases. J.R.W. and W.C.N. provided the sequence and annotation for *A. fumigatus*. T.T., T.K., K.A. and M. Machida provided the sequence and annotation for *A. oryzae*. D.W.D. co-ordinated interactions between the different sequencing centres and scientific communities. J.E.G. and S.A.O. coordinated the comparative analyses. J.E.G. wrote and edited the paper, and produced the figures.

Author Information The *A. nidulans* genome sequence has been deposited at DDBJ/EMBL/GenBank under the project accession AACD00000000. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests.
Correspondence and requests for materials should be addressed to J.E.G. (jgalag@mit.edu).

© 2005 Nature Publishing Group